

```

1 TCACAGAACA TGTCCAACAA CAGCCCCGAG TATGCTTTGG TTTTCACCAT
51 CTCGGGTGCT ATGGCCACCA TGGTCTCCAG TGGCCTGGGT GCTGCCTGTG
101 GCATGGCCAA GAATGGCACC GGCATCATGG CCATGTCTGT CATGTGGCCA
151 GAGCTGATCC ACATGAAGTC CATCATCCCA GTGGTCATGG CTGGTATCAT
201 CACCATCTAT GGCCTAGTGG CGGCTGTCCC CCCTGCCAAC TCCCTGAATG
251 ATGACAACAG TCTCTATAGC AGTTTCCTCC AGCTGGGCGC TGGCCTGAGT
301 GGCCTGGCAG CCGGCTTTGC CATCGTCATC GTGGGGGACA CTGGCAAGTG
351 TGGCACTGCC CAGCAGCCCC GACTATTTGT AGGCATGATA CTGATCCTCA
401 TCTTGCCAA GGTGCTCATT CTCTCCACAA AGCAGCCCCT CTCAAAACCC
451 ACCAGTCACA GAATACGATG TAAAGACCAC CCCTCCTCAT TCCGGAACAA
501 ACAGCCTGAC ACGCATGTGC TGGGCAGCTG GCCCTCAGTA GTTGATCTTC
551 TAAGTGATACA GTGTCCTCGT GTTCATCGTC TGTGCGCCAG GCCTTGCCCC
601 CTCCCGCCCC ATGCTGTGGA CATCTGAACC TAC

```

**FEATURES:**

```

5'UTR:      1-9
Start Codon: 10
Stop Codon: 625
3'UTR:      628

```

# **HOMOLOGOUS PROTEINS:**

## Top BLAST Hits:

	Score	E
gi 4502313 ref NP_001685.1  ATPase, H+ transporting, lysosomal ...	181	5e-45
gi 137477 sp P23956 VATL_BOVIN VACUOLAR ATP SYNTHASE 16 KD PROT...	180	8e-45
gi 227919 prf  1713409A H ATPase 16K [Bos taurus]	180	1e-44
gi 3024812 sp O18882 VATL_SHEEP VACUOLAR ATP SYNTHASE 16 KD PRO...	178	3e-44
gi 418179 sp Q03105 VATL_TORMA VACUOLAR ATP SYNTHASE 16 KD PROT...	175	5e-43
gi 6753144 ref NP_033859.1  ATPase-like vacuolar proton channel...	174	8e-43
gi 67954 pir  PXBOV6 H+-transporting ATPase (EC 3.6.1.35), vacu...	165	3e-40
gi 137478 sp P23380 VATL_DROME VACUOLAR ATP SYNTHASE 16 KDA PRO...	157	1e-37
gi 3334403 sp O16110 VATL_AEDAE VACUOLAR ATP SYNTHASE 16 KD PRO...	156	2e-37
gi 1718095 sp P55277 VATL_HELVI VACUOLAR ATP SYNTHASE 16 KD PRO...	153	1e-36
gi 401334 sp P31403 VATL_MANSE VACUOLAR ATP SYNTHASE 16 KD PROT...	153	1e-36
gi 10442628 gb AAG17394.1 AF277150_1 (AF277150) V-ATPase 16 kD ...	151	5e-36
gi 7294725 gb AAF50062.1  (AE003544) CG7547 gene product [Droso...	151	7e-36
gi 2493142 sp Q26250 VATL_NEPNO VACUOLAR ATP SYNTHASE 16 KD PRO...	150	2e-35
gi 251354 gb AAB22509.1  vacuolar H(+)-ATPase proteolipid subun...	150	2e-35
gi 2493143 sp Q00607 VATL_CANTR VACUOLAR ATP SYNTHASE 16 KD PRO...	144	9e-34

## BLAST to dbEST:

	Score	E
gi 9336427 /dataset=dbest /taxon=960...	165	1e-38
gi 6359805 /dataset=dbest /taxon=9606 ...	165	1e-38
gi 9134224 /dataset=dbest /taxon=9606...	165	1e-38
gi 10219114 /dataset=dbest /taxon=96...	165	1e-38
gi 9347217 /dataset=dbest /taxon=960...	165	1e-38
gi 9152104 /dataset=dbest /taxon=9606...	161	2e-37
gi 9894156 /dataset=dbest /taxon=960...	155	1e-35

## **EXPRESSION INFORMATION FOR MODULATORY USE:**

### library source:

### Expression information from BLAST dbEST hits:

gi|9336427 Human uterus  
gi|6359805 Human fetal liver  
gi|9134224 Human brain  
gi|10219114 Human lung  
gi|9347217 Human placenta  
gi|9152104 Human skin  
gi|9894156 Human ovary

### Expression information from PCR-based tissue screening panels

Human Bone marrow  
Human Brain  
Human Colon  
Human Fetal Brain  
Human fetal heart  
Human Fetal Kidney  
Human fetal liver  
Human Heart  
Human Kidney  
Human Liver  
Human Lung  
Human Pancreas  
Human Placenta  
Human Prostate  
Human Skeletal Muscle  
Human Small Intestine  
Human Spleen  
Human Testis

```

1 MSNNSPEYAL VFTISGAMAT MVSSGLGAAC GMAKNGTGIM AMSVMWPELI
51 HMKSIIIPVVM AGIITIYGLV AAVPPANSLN DDNSLYSSFL QLGAGLSGLA
101 AGFAIVIVGD TGKCGTAQQP RLFVGMILIL IFAKVLILST KQPLSKPTSH
151 RIRCKDHPSS FRNKQPDTHV LGSWPSVVDL LSVQCPRVHR LLARPCPLPP
201 HAVDI

```

# **FEATURES:**

## **Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATIONN-glycosylation site  
35-38 NGTG

[2] PDOC00005 PS00005 PKC\_PHOSPHO\_SITEProtein kinase C phosphorylation site

Number of matches: 4

```

1      111-113 TGK
2      139-141 STK
3      149-151 SHR
4      160-162 SFR

```

[3] PDOC00006 PS00006 CK2\_PHOSPHO\_SITECasein kinase II phosphorylation site

Number of matches: 2

```

1      78-81 SLND
2      176-179 SVVD

```

[4] PDOC00008 PS00008 MYRISTYLN-myristoylation site

Number of matches: 8

```

1      16-21 GAMATM
2      25-30 GLGAAC
3      27-32 GAACGM
4      31-36 GMAKNG
5      68-73 GLVAAV
6      93-98 GAGLSG
7      98-103 GLAAGF
8      172-177 GSWPSV

```

#### Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	14	34	1.889	Certain
2	37	57	0.733	Putative
3	60	80	2.030	Certain
4	95	115	1.775	Certain
5	127	147	1.699	Certain

#### BLAST Alignment to Top Hit:

```
>gi|4502313|ref|NP_001685.1| ATPase, H+ transporting, lysosomal
(vacuolar proton pump) 16kD
>gi|137479|sp|P27449|VATL_HUMAN VACUOLAR ATP SYNTHASE 16
KD PROTEOLIPID SUBUNIT >gi|107394|pir||A39367
H+-transporting ATPase (EC 3.6.1.35) chain PKD1 - human
>gi|189676|gb|AAA60039.1| (M62762) vacuolar H+ ATPase
proton channel subunit [Homo sapiens]
Length = 155
```

Score = 181 bits (455), Expect = 5e-45

Identities = 110/153 (71%), Positives = 114/153 (73%), Gaps = 14/153 (9%)

```
Query: 2   SNNSPEYALVFTISGAMATMVSSGLGAACGMAKNGTGIMAMSVMWPELIHMKSIIPVVMA 61
          S + PEYA  F + GA A MV S LGAA G AK+GTGI AMSVM PE I MKSIIPVVMA
Sbjct: 4   SKSGPEYASFFAVMGASAAMVFSALGAAYGTAKSGTGIAAMSVMRPEQI-MKSIIPVVMA 62

Query: 62  GIITIYGLVAAPPPANSLNDDNSLYSSFLQLGA----GLSGLAAGFAIVIVGDTGKCGTA 117
          GII IYGLV AV  ANSLNDD SLY SFLQLGA  GLSGLAAGFAI IVGD G  GTA
Sbjct: 63  GIITAIYGLVVAVLIANSLNDDISLYKSFLQLGAGLSVGLSGLAAGFAIGIVGDAGVRGTA 122

Query: 118 QQPRLFVGMILILIFAKV-----LILSTK 141
          QQPRLFVGMILILIFA+V             LILSTK
Sbjct: 123 QQPRLFVGMILILIFAEVLGLYGLIVALILSTK 155
```

#### Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF00137	ATP synthase subunit C	14.8	0.028	2

#### Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00137	1/2	12	73	..	1 65	7.6	2.4
PF00137	2/2	89	133	..	1 53	14.6	0.031

1 GCTGTGGGGC CAGGAAAAGG AGAGAAGGTG AAACCCCGT CAGTCCCTCA  
 51 CAATCAGCAC GTGGAAATCT AGAAATGCAG GAGAGGCGTG GACTCATGGT  
 101 GGAATCCAGA ATGAAAGAGG TGGACGACTG AATGAGCAGA AGGAGGCAAG  
 151 CACCAGAGGC TTGGGGGTCA CATTCTTGG AAGTGGCCTG GAGCTGGCAG  
 201 ATGAGAACTC TGGCTACCCG TCCTCATTC ACTAACAGTA GCTCCTCTAA  
 251 CGACATGCCC CTTCCCTCTG TACCCCGCTC CGCATGCGGC AAGTAGTTCC  
 301 CGGACGCGAC CCTTCCCCCT GTACCCCGCT CCGCATGGGG CCAGTAGTTC  
 351 CCGGACGCGC CCCTTCCCTC TGTACCCCGC TCCGCATGCG GCAAGTAGTT  
 401 CCGGACGCGC CCCCTTCCCT CTGTACCCGG CTCGCGATGC GGAAAGTAGT  
 451 TCCTACGGTG TTGGTTTTGC ATGTAGATGA AACCTTTGA GGGGTAAAGG  
 501 TTTTTTTTTT AAGTACTTTA GCAAATGCAA ACTGTTATTA TCAATATTAG  
 551 CCAGCATCTT TTTTTTTTTT TTTTTTTTTT TTTTGTAGAT GGAGTTTCGC  
 601 TCTTGTACCC CAGGCTGGAG TGCAATGGCA AAATCTAGGC TCACTGCAAC  
 651 CTCGCGCTCC CAAGTTCAAG CGATTCTCCT GCCTCAGCCT CCCAGGTAGC  
 701 TGGGATTACA GCGGTGTGCA ACCACACCCA GCTAATTTT GTATTTTATG  
 751 TAGAGACAGG GTTTCACCAT GTTGGCCAGG CTGGTCTCGA ACTCCTGACC  
 801 TCATGTGATC CATCCGCCTC AGCCTCCCAA AGTGCTGGGA TTACGTAGCC  
 851 AGTGTCTTTC TTAAGTGCCT GTCAAATAAT GCTCCTGGTT TATAAGTGCC  
 901 CCTGGCTCTA CCTTCTGGGT GCTCAGACAC CAACACAGAG AGAACAGAAT  
 951 TAACATCCTG AGAAGTTACA TATGCTAAAA TATAAAGAGT AAGATTGTGA  
 1001 GGAAACTGCA GGGGAAGCAG GTAGGTTAGG AAAAGGTATC CTCACTTTTC  
 1051 TGCTGACCGA TGAGTCATAA TTCTTGAATT TCGGTGCTGG AAAGGTCCAT  
 1101 TAAGCATTC AGGAGATTCT AGGGAGCTTC CAGAATGGTA GAAGAAGTGG  
 1151 AACCATAAAG CCTGGGGAAG GGATGGAAGT CCTTGGGAAA GAAGCACTAA  
 1201 ACAGCCAGT GGAGACAAGG AAGGACTGGT CTCTCCTGTG CTTGAGCCCC  
 1251 AGCAATGATT ATTCACTCAG ATATGCCCCG GCAGGTCCCTG CTGCTAGAGC  
 1301 CAGTGCTGTT CCCAGACCCA GGCAAGGTGC CATCCTACCC CTGACAGGAA  
 1351 ACAGGCGAGG AGGTGGGGCT GCCCGGGTG CCTGGTGTG GGAGGGGCGG  
 1401 GGGGAATCC CGGTGTGGT GAGGACAAGG CAGAGTCAGC TAGCTGTGAG  
 1451 GCTAGGGGAG AAGACCTCTC TAGTCTGGGA GAGACCCCTC CTTTCCTAGC  
 1501 TCCTTGTACT TCCAAAAGG CAGGCTTCCT GCTGTTACTA ACCATACCAG  
 1551 GACTGACTAT ACAGCAGCCA GAAAGATTCT GAGAAACCTG TGATAGAGAA  
 1601 AAACAGATGC GGAAGCGGGA GAAGAGAATT TCATAGGACA CTAGGGAAAG  
 1651 AGAATGGGAA CTTGTGGTCT AAAGAGGGAA CCAAGTCTGG CCAACATGGT  
 1701 GAAACCCCAT CTTATTACA AATACAAAAA TTAGCTGGGC ATGGTAGTGC  
 1751 ATGCCTGTAA TCCCAGCTAC TCAGGAGGGT AAGGCATGAG AATCACTTGA  
 1801 GACTGGGAGG CGGAGGTTGC AGTGAGCCGA GACTGCACCA CTGCACTCCA  
 1851 GCCTGGGCAA CAGAGCAAGA CCTCGTCTCA AAAAAAAAAA AAAATTAAAT  
 1901 TTAAATTAAA AAAAAATAAA CAGGGAACCA ACAAGAGCTG GCAGAACAGA  
 1951 ATAAAGTCTC AAGCCAAATA ACTCCCTTGC CTTGGAAGAA CAAGGCTGCC  
 2001 AGCGCTCTTG AGCCTCTGTT TATCGGTAC CAGTTCAAAG GACAGTGAGC  
 2051 CTGAGCTGGC CTGGGAGGCC CTCGCCCTC CCAGATGAAA ACAATAGGCC  
 2101 TGTTTTCTCT AGCTCTTCTT GTAATCCAGA AGGCCACAC AGAGAGGAAG  
 2151 AGGGGGGCAA AGGCAGTGGC TATACCCAGT GGGGGAGGGG ATATTTAGCC  
 2201 TCCCATAAAT TCATCAGCTC CCTTAAAGAC ACCCCAAAAC ACCAACATC  
 2251 TAAGTGTTAA AATAGTGAAT GCTATGCAAA TGGAGCTTTA AAACCTATCC  
 2301 CTTAGCCCCA GTACCACCAG ATTACTAACC CTAAACCCCA TCTGTAGGAG  
 2351 ATATTCTGAA GCCACCACAG GGAAGGGAT AAGGGCCTGA GAGACAAAGG  
 2401 ACAGATGGGG TCTCCCAAC AATTAAAGTT AAGTTCCACA AGGATACAGT  
 2451 ACTGGCAGAG ATTTGGAAGT AGGGGCAAGT ATTCTGACAG AAGGGTGGTG  
 2501 TCTTAGGCAC CCTTCAATTA GGAGTAGCTA AAGGCTGTGT GTGTGTCTGT  
 2551 GTGTGTGCAT AAGAAAAGAA ATAGGAGGGT GTGTGTGTGG TAAGAAAGAG  
 2601 CATCTTGGCT GGGCGCGGTG GCTCACCCCT ATAATCGCAG CACTTTGGGA  
 2651 TGCCAAGGCT GCGGATTGC CTGAGCTCAG GAGTTTGAGA CCATACGGGG  
 2701 CAACATGGTG AAACCCATC TCTACTAAAA ATACAAAAAA TTAGCTGGGC  
 2751 ATGGTGGTGC GTGCCTATAG TTCCAGCTAC TCGGGAGGCT GAGGCATGAG  
 2801 AATGGCTTGA GCCCTGGAGG CAGAGGTTGA AGTGAGCTGA GATCGCACCA  
 2851 TTGCATTCCA GCTTGGGCTA CAGAGTGACA CTCCATCTCA AAAAAAAAAA  
 2901 AAAAAAAAAA AAAAACAGC ATCTTTGCTG CCACTAGTCC ACTGTCTTTG  
 2951 CACTCACTCT CTGCCATGCC CATCCTTGTC CCCCTCCCCA CTCACAGACA  
 3001 TGTCCAAACA CAGCCCGAG TATGCTTTGG TTTTACCAT CTCGGGTGCT  
 3051 ATGGCCACCA TGGTCTCCAG TGGCCTGGGT GCTGCCTGTG GCATGGCCAA  
 3101 GAATGGCACC GGCATCATGG CCATGTCTGT CATGTGGCCA GAGCTGATCC

FIGURE 3, page 1 of 4

3151	ACATGAAGTC	CATCATCCCA	GTGGTCATGG	CTGGTATCAT	CACCATCTAT
3201	GGCCTAGTGG	CGGCTGTCCC	CCCTGCCAAC	TCCCTGAATG	ATGACAACAG
3251	TCTCTATAGC	AGTTTCTCTC	AGCTGGGCGC	TGGCCTGAGT	GGCCTGGCAG
3301	CCGGCTTTGC	CATCGTCATC	GTGGGGGACA	CTGGCAAGTG	TGGCACTGCC
3351	CAGCAGCCCC	GACTATTTGT	AGGCATGATA	CTGATCCTCA	TCTTTGCCAA
3401	GGTGCTCATT	CTCTCCACAA	AGCAGCCCCT	CTCAAAACCC	ACCAGTCACA
3451	GAATACGATG	TAAAGACCAC	CCCTCCTCAT	TCCGGAACAA	ACAGCCTGAC
3501	ACGCATGTGC	TGGGCAGCTG	GCCCTCAGTA	GTTGATCTTC	TAAGTGATACA
3551	GTGTCTCGT	GTCATCGTC	TGTTGGCCAG	GCCTTGCCCC	CTCCCGCCCC
3601	ATGCTGTGGA	CATCTGAACC	TACTCATCAC	CCATCCAGGT	CCCCGACCAG
3651	TGAGGACTCA	GGCCCCTGGA	TGCCCCACCC	ATCTCCCTTG	AGTACTCTAT
3701	GTATAAGGAT	GAATTAGAGT	TGTCATTTTC	TCTTCATTAG	ATATTTATAA
3751	AGATTTGGCC	TGTCCATACC	CCTGTGGAGC	AGCCCTCATC	TCCCACCTAT
3801	CTGTCACGTC	ATGGAGGTTT	CCATTGCGGA	GGCTCCTTGG	ATGGAACCAC
3851	CCTCTCCAGC	CCGCGCTGCC	AGGCCCTGTG	CGGCAGCTGT	GTCTGATAAA
3901	GTTCTCAGAT	GTGCGGGGAG	GGAAAGAAAA	AAAAAAGAGA	GTGTGAGTAC
3951	GTAAGAGAGA	GAAAACGGGA	GTGGGTGTGT	GAGCTGGAGA	CAGGGAAGTG
4001	GCAGGAAAAG	TCTGATAAGA	TCACCTCCTT	CCTACCCAAG	CAGAGATACT
4051	GGACACAGCC	CCTCAAGGAC	CCAGAGGGTA	AGTAGAGCGC	GAGATGCTTG
4101	CCTTTCTCAA	TGGGAGGTGG	CCTCCCAGGC	CTGAAGAAGT	CTCCATTTAC
4151	CCCAGAGCCA	ACTAGGAAGC	AGGTAGACAG	CATCATCCCC	ACTTATACCC
4201	CAAGGTGCTT	GGGTGAATG	GCAGGCCCAA	AGCCAAAGCA	TGAGACAGAT
4251	TAAATGTTTC	TATGGCGAGA	GAAGGAGAAG	GGGTCAACAG	CATCTCTCCA
4301	CTGAGCAAAT	GAAAGGAAGA	GAGAAGGCAG	GCTGATACCC	TCATCAATTT
4351	CCTACTGTCC	ATGATATAAC	ACCATCAACT	GGACTTTTTT	TTTTTTTTTG
4401	AGATAGAGTC	TCGCTTTTGT	CACCCAGGCT	GGAGTGCAGT	GGCATGATCT
4451	CAGCTCACTG	CAACTTCCAT	CTCCCAGGTT	CAAGTGATTG	TCCCGCCTCA
4501	GCCTCCTGAG	TAGCTGGGAT	TACAGGTGCC	TGCTACCACA	TCCAGCTGAT
4551	TTTTTTTGTA	TTTTTAGTAG	AGATGGGGTT	TCTTTCTTTT	TTTTTTTTTT
4601	TTTTGAGACG	GAGTCTTGCT	CTGTCGCCCC	GGCTGGAGTG	CAGTGGCGCG
4651	ATCTCGGCTC	ACTGCAACCT	CCGCTCCCCA	GGTTCACGCC	ATTCTCCTGC
4701	CTCAGCCTCC	CGAGTAGCTG	GGACTACAGG	CACCTGCCAC	CACACTCGGC
4751	TAATTTTTTG	TATGTTTAGT	AGATATGGGG	TTTCACTGCT	GTCTCAACCT
4801	TCTGACCTCA	TGATCCGCC	GCCTCGGCCT	CCCAAAGTGC	TGGGATTACA
4851	GGCATGAGCC	ACTGTGCCCC	GCCTTTTTTT	TTTTTTTTGA	GATGGAGTCT
4901	CGCTCTGTCT	CCCAGGCTGG	AGTGCAATGC	CACAATCTCA	GCTCACTGCA
4951	AGCTCCACCT	CCGAGGTTCA	CGCCATTCTC	CTGCCTCAGC	CTCCTGAGTA
5001	GCTGGGACTA	CAGCGCCCCG	CCACCACGCC	CAGCTAATTT	TTTGTATTTT
5051	TAGTAGAGAC	GGGGTTTCAC	CTTGTTAGCC	AGGATGGTCT	TGATCTCCTG
5101	ACCTCGTGAT	CCACCTGCCT	CAGCCTCCCA	AAGTGCTGGG	ATTACAGGTG
5151	TGAGCCACCA	TGCCTGGCCT	TTTTTTTTTT	TTTAAGACAG	GAGTGTGGTG
5201	GCACAATCTC	AGCTCACTGC	AACCTCCCCT	TCTAGGTTCA	AGCAATTCTC
5251	CTGCCTCAGC	TTCTTAAGTA	TAGTAATAGC	TGGGACTATA	GGCGCCCACC
5301	ACCACGCCCC	GCTAATCTTT	TGTATTTTTA	GTAGAGATGG	GGTTTCACCA
5351	TGTTGGCCAG	GCTGGTCTCG	AATTGCTGAC	CTCAAGTGAT	CTGCCACCT
5401	GGGCCTCCCA	AAGTGCTGGG	ACTATAGGCG	GGAGCCACCG	CGCCACGCTT
5451	GGACTCTTTT	TAATGAAGCC	TTCAAAAAAA	CTCCTTTTCT	CAGCGCTTCT
5501	TACTCTCTGA	AACAGACTCT	CCACTCTGCT	AACCCTGCCT	CTCACACTGT
5551	GGAACCTAAC	CGGATCTTTT	TATTCTGAAT	CCACAACGTG	AAGTACTTGT
5601	CCTCTGTCTA	TCGATGGCTA	CCTGTGTTTT	GAAGTGTTTT	TATGGGAATG
5651	AAGCACTGGA	GGGGAGGAAA	TCAGGCCAGT	TCTAGAAGTA	GAAGGAAGGC
5701	GAAGAAACCA	GGAAAAATAT	TTATGTGATG	GGAGGAAAGG	CAGTTTATAA
5751	ATCACTCATG	GATCTCTATG	CCAGAGGGAT	GTGTGAGACA	CACGCATGCA
5801	CACACACACT	GACTTGCAGG	TACATGCAGA	GGCAGAAACA	AGTCAGGACA
5851	TGACACATAC	ATGAATACAC	ATACCATTCT	CATCAGAAAC	CAGTCAGAGC
5901	AGAGGGGCCC	TGCCTGGAGC	AAGGAGACTG	GAATTTATTC	CCCTCCTCCT
5951	CTCAAAGGGT	AATTTTGCTG	CCTCCATGTC	TAGGTTCCCC	ACAGATCTGG
6001	CTGCCTCAGA	CAGGGGCCCT	GGTCTGGTGG	CTGGACTCAG	CCTGGAGGTC
6051	TTACACAGATG	GAGGCCTATA	AGAGGTGGCA	GCTGACACCT	GGAGGGAGCT
6101	GGATGAAAGC	AGGCAGTGCA	GAGTAGAGAA	AGCCAGGTGG	TGGGGGAGGG
6151	AGTGAGGGAG	AAGAGGGGAC	CAGATTCAAG	CAGCCTTGCG	CTGGTTCTAA
6201	AATGGCCACA	GCAAGGCAAC	GGACAGATGG	TCCCTTTCTG	ATGCTGAGCC
6251	GGGGAAGTGG	GGAAAGGGAA	AAGGAAAAAA	TAAACACCAT	CACAGTCAGA

FIGURE 3, page 2 of 4

6301 AATTTAAAA TAAACTGAAA AACCTAAAA ATAAACCGT

**FEATURES:**

Start: 3000  
Exon: 3000-3614  
Stop: 3615

**CHROMOSOME MAP POSITION:**

Bac accession number: AC005973  
Chromosome 17

**ALLELIC VARIANTS (SNPs):**

DNA			
Position	Major	Minor	Domain
559	-	A T	Beyond ORF (5')
3638	G	C	Beyond ORF (3')
5446	C	T A	Beyond ORF (3')
5808	A	G	Beyond ORF (3')
5892	A	C	Beyond ORF (3')
6071	A	G	Beyond ORF (3')

**Context:**

**DNA**

**Position**

559 CCCTTCCCTCTGTACCCCGCTCCGCATGCGGCAAGTAGTTCCCGGACGCGACCCCTTCCCC  
CTGTACCCCGCTCCGCATGGGGCCAGTAGTTCCCGGACGCGCCCTTCCCTCTGTACCCCG  
GCTCCGCATGCGGCAAGTAGTTCCCGGACGCGCCCTTCCCTCTGTACCCGGCTCCGCAT  
GCGGAAAGTAGTTCCCTACGGTGTGGTTTGCATGTAGATGAAACCCCTTTGAGGGGTAAA  
GGTTTTTTTTTTAAGTACTTTAGCAAATGCAAAGTATTATTATCAATATTAGCCAGCATC  
[- , A, T]  
TTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTTTCGCTCTTGTCACCCAGGCTGGA  
GTGCAATGGCAAAATCTAGGCTCACTGCAACCTCCGCTCCCAAGTTCAAGCGATTCTCC  
TGCCTCAGCTCCAGGTAGCTGGGATTACAGGCGTGTGCAACCACACCCAGCTAATTTT  
TGTATTTTGTAGTAGAGACAGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGAC  
CTCATGTGATCCATCCGCTCAGCTCCCAAGTGTGGGATTACGTAGCCAGTGTCTTT

3638 GTGTGGCACTGCCCAGCAGCCCCGACTATTTGTAGGCATGATACTGATCCTCATCTTTGC  
CAAGGTGCTCATTCTCCACAAAGCAGCCCTCTCAAAACCCACAGTCACAGAATACG  
ATGTAAAGACCACCCCTCCTCATTCCGGAACAAACAGCCTGACACGCATGTGCTGGGCAG  
CTGGCCCTCAGTAGTTGATCTTCTAAGTGTACAGTGTCTCGTGTTCATCGTCTGTTGGC  
CAGGCCTTGCCCCCTCCGCCCCATGCTGTGGACATCTGAACCTACTCATCACCATCCA  
[G, C]  
GTCCCCGACCAGTGAGGACTCAGGCCCTGGATGCCCCACCCATCTCCCTTGAGTACTCT  
ATGTATAAGGATGAATTAGAGTTGTCTTTCTCTTCATTAGATATTTATAAAGATTTGG  
CCTGTCCATACCCCTGTGGAGCAGCCCTCATCTCCACCTATCTGTCACGTCATGGAGGT  
TCCCATTGCGGAGGCTCCTTGGATGGAACACCCCTCTCCAGCCCGCGCTGCCAGGCCCTG  
TGCGGCAGCTGTGTCTGATAAAGTTCTCAGATGTCGCGGGAGGGAAAGAAAAAAGAA

5446 AGGTGTGAGCCACCATGCCTGGCCTTTTTTTTTTTTTTTAAGACAGGAGTGTGGTGGCACA  
ATCTCAGCTCACTGCAACCTCCCCTTCTAGGTTCAAGCAATTCTCCTGCCTCAGCTTCTCT  
AAGTATAGTAATAGCTGGGACTATAGGCGCCACACACGCCCCGCTAATCTTTTGTAT  
TTTTAGTAGAGATGGGGTTTACCATGTTGGCCAGGCTGGTCTCGAATTGCTGACCTCAA  
GTGATCTGCCACCTGGGCCTCCCAAAGTGTGGGACTATAGGCGGGAGCCACCGCGCCC  
[C, T, A]  
GCCTGGACTCTTTTAAATGAAGCCTTCAAAAAAACTCCTTTTCTCAGCGCTTCTTACTCT  
CTGAAACAGACTCTCCACTCTGCTAACCTGCCTCTCACACTGTGGAACCAACCGGATC  
TTTTTATTCTGAATCCACAACGTGAAGTACTTGTCTCTGTCTATCGATGGCTACCTGTG  
TTTTGAAGTGTTTTTATGGGAATGAAGCACTGGAGGGAGGAAATCAGGCCAGTTCTAGA  
AGTAGAAGGAAGGCGAAGAAACAGGAAAAATATTTATGTGATGGGAGGAAAGGCAGTTT

5808 TGAACAGACTCTCCACTCTGCTAACCTGCCTCTCACACTGTGGAACCAACCGGATCT  
TTTTATTCTGAATCCACAACGTGAAGTACTTGTCTCTGTCTATCGATGGCTACCTGTGT  
TTTGAAGTGTTTTTATGGGAATGAAGCACTGGAGGGGAGGAAATCAGGCCAGTTCTAGAA  
GTAGAAGGAAGGCGAAGAAACCAGGAAAAATATTTATGTGATGGGAGGAAAGGCAGTTTA  
TAAATCACTCATGGATCTCTATGCCAGAGGGATGTGTGAGACACACGCATGCACACAC  
[A, G]  
CTGACTTGCAGGTACATGCAGAGGCAGAAACAAGTCAGGACATGACACATACATGAATAC  
ACATACCATTCTCATCAGAAACCAGTCAGAGCAGAGGGGCCCTGCCTGGAGCAAGGAGAC  
TGGAATTTATTTCCCTCCTCCTCTCAAAGGGTAATTTTGCTGCCTCCATGTCTAGGTTCC  
CCACAGATCTGGCTGCCTCAGACAGGGGCCCTGGTCTGGTGGCTGGACTCAGCCTGGAGG  
TCTTCACAGATGGAGGCCTATAAGAGGTGGCAGCTGACACCTGGAGGGAGCTGGATGAAA

5892 AGTACTTGTCTCTGTCTATCGATGGCTACCTGTGTTTTGAAGTGTTTTTATGGGAATGA  
AGCACTGGAGGGGAGGAAATCAGGCCAGTTCTAGAAGTAGAAGGAAGGCGAAGAAACCAG  
GAAAAATATTTATGTGATGGGAGGAAAGGCAGTTTATAAATCACTCATGGATCTCTATGC  
CAGAGGGATGTGTGAGACACACGCATGCACACACACACTGACTTGCAGGTACATGCAGAG  
GCAGAAACAAGTCAGGACATGACACATACATGAATACACATACCATTCTCATCAGAAACC  
[A, C]  
GTCAGAGCAGAGGGGCCCTGCCTGGAGCAAGGAGACTGGAATTTATTTCCCTCCTCCTCT  
CAAAGGGTAATTTTGCTGCCTCCATGTCTAGGTTCCCCACAGATCTGGCTGCCTCAGACA  
GGGGCCCTGGTCTGGTGGCTGGACTCAGCCTGGAGGTCTTCACAGATGGAGGCCTATAAG  
AGGTGGCAGCTGACACCTGGAGGGAGCTGGATGAAAGCAGGCAGTGCAGAGTAGAGAAAG  
CCAGGTGGTGGGGAGGGAGTGAGGGAGAAGAGGGGACCAGATTCAAGCAGCCTTGCGCT

6071 CCAGAGGGATGTGTGAGACACACGCATGCACACACACACTGACTTGCAGGTACATGCAGA  
GGCAGAAACAAGTCAGGACATGACACATACATGAATACACATACCATTCTCATCAGAAAC  
CAGTCAGAGCAGAGGGGCCCTGCCTGGAGCAAGGAGACTGGAATTTATTTCCCTCCTCCT  
CTCAAAGGGTAATTTTGCTGCCTCCATGTCTAGGTTCCCCACAGATCTGGCTGCCTCAGA  
CAGGGGCCCTGGTCTGGTGGCTGGACTCAGCCTGGAGGTCTTCACAGATGGAGGCCTATA  
[A, G]  
GAGGTGGCAGCTGACACCTGGAGGGAGCTGGATGAAAGCAGGCAGTGCAGAGTAGAGAAA  
GCCAGGTGGTGGGGGAGGGAGTGAGGGAGAAGAGGGGACCAGATTCAAGCAGCCTTGCGC  
TGGTTCTAAAATGGCCACAGCAAGGCAACGGACAGATGGTCCCTTTCTGATGCTGAGCCG  
GGGAAGTGGGGAAAGGGAAAAAGGAAAAATAAACACCATCACAGTCAGAAATTTAAAAAT  
AAACTGAAAAACCTAAAAAATAAACCGT